# Measures of Disease-Exposure Association


- Relative risk RR
- Odds ratio OR
- The OR as an approximation to the RR
- Symmetry of roles of disease and exposure in the OR
- Excess risk ER
- Attributable risk AR

# Measures of Disease-Exposure Association

Purposes of epidemiology:

- **quantification** of the occurrence of a disease [*descriptive* studies]
- **strength of the association*** between exposure and the onset of the event [*analytical* studies]

Estimate measures of *disease-exposure* association (or **measures of effect**).

**Disease frequency** in the **exposed group** is compared with the frequency of disease in the **group of those not exposed**, making use of the appropriate measure of occurrence.

This comparison can occur in two ways: in **absolute** terms and in **relative** terms.

***note that we are not using the term *causal* effect!**

In a *general* sense, each disease is the effect of one (or more…) causes.

In a *quantitative* sense, an **effect** is the measure of **diversity in the occurrence** of a pathology in two [or more..] groups that differ by *one certain* feature [*univariable* analysis].

- absolute scale: **difference** between two prevalences, two risks (Cum Inc) or two incidence rates
- relative scale: **ratio** of two prevalences, two risks (Cum Inc) or two incidence rates
- attributable risk: **proportion of cases** *attributable* to exposure in a population

Does a mother's marital status affect the risk of a baby's death in the first year? ***To what extent***? What about birthweight?

# Relative risk

The Relative Risk for an outcome D associated with a *binary* risk factor E, denoted by RR, is defined as follows:

$$RR = \frac{P(D|E)}{P(D|\bar{E})}$$

| | D | Not D | Tot |
|---|---|---|---|
| E | a | b | a+b |
| Not E | c | d | c+d |
| Tot | a+c | b+d | N |

UNITÀ DI BIOSTATISTICA
Dipartimento Universitario Clinico di Scienze Mediche Chirurgiche e della Salute

Some simple implications immediately follow:

RR
- <1 : lower risk or probability of *D* when exposed than when unexposed
- =1 : *null* value equivalent to saying that *D* and *E* are independent $P(D|E) = P(D|\bar{E})$
- >1 : greater risk or probability of *D* when exposed than when unexposed

The Relative Risk is the basis of a *multiplicative model for risk* :

$$Risk_{Exposed} = Risk_{unexposed} * RR$$

*Baseline* Risk

If you smoke cigarettes, your lifetime risk of lung cancer increases tenfold, i.e., the Relative Risk for lung cancer associated with cigarette smoking is 10.


UNITÀ DI BIOSTATISTICA
Dipartimento Universitario Clinico di Scienze Mediche Chirurgiche e della Salute

Restrictions on the range:

$$0 < RR \leq \frac{1}{P(D|\bar{E})}$$

For instance, if $P(D|\bar{E}) = 1/3$ (30%) then RR $\leq$ 3 since $P(D|E) \leq 1$

This *restriction* could become an issue with *common* diseases

RR is ***not symmetric*** in the role of the two factors D and E.

The Relative Risk for E associated with D is a different measure of association:

$$\frac{P(D|E)}{P(D|\bar{E})} \neq \frac{P(E|D)}{P(E|\bar{D})}$$

## Relative Effect

| Infant Mortality | Mother's Marital Status: Unmarried | Married |
|---|---|---|
| Death | 16,712 | 18,784 |
| Live at 1 year | 1,197,142 | 2,878,421 |
| Total | 1,213,854 | 2,897,205 |

The Relative Risk for infant mortality in the U.S. in 1991, associated with a mother being unmarried at the time of birth, is:

$$RR = \frac{16712}{1213854} : \frac{18784}{2897205} = 2.12$$

the risk of an infant death with an unmarried mother is **double** the risk w.r.t. mother is married.

UNITÀ DI BIOSTATISTICA
Dipartimento Universitario Clinico di Scienze Mediche Chirurgiche e della Salute

## Relative Effect

The RR for infant mortality in the U.S. in 1991, associated with a low-birthweight infant, is:

| | Birthweight | | |
|---|---|---|---|
| Infant Mortality | Low Birthweight | Normal Birthweight | Total |
| Death | 21,054 | 14,442 | 35,496 |
| Live at 1 year | 271,269 | 3,804,294 | 4,075,563 |
| Total | 292,323 | 3,818,736 | 4,111,059 |

$$RR = \frac{21054}{292323} : \frac{14442}{3818736} = 19.0$$

Much greater effect of birthweight on infant mortality than we saw for a mother's marital status.

# Odds Ratio

An alternative quantity that is used is the ***odds*** of D as given by: $\dfrac{P(D)}{P(\bar{D})}$

The odds gives the likelihood of D occurring relative to it not occurring: "how likely am I to win?" as compared to "how likely am I to lose?"

An even odds event D (odds of D are 1) is equivalent to P(D)=1/2, that is, the **same chance** of winning as losing.

$$Odds = \frac{Probability\ of\ event}{1 - Probability\ of\ event}$$

The Odds Ratio measures association by comparing the odds of D in the exposed and unexposed.

The Odds Ratio for D associated with E is defined by:

$$OR = \frac{P(D|E)}{P(\bar{D}|E)} : \frac{P(D|\bar{E})}{P(\bar{D}|\bar{E})} \qquad OR = \frac{P_{exp}/(1 - P_{exp})}{P_{unexp}/(1 - P_{unexp})}$$

OR:

<1 : lower risk or probability of *D* when exposed

=1 : *null* value equivalent to saying that *D* and *E* are independent

>1 : greater risk or probability of *D* when exposed

The Odds Ratio is also the basis of a *multiplicative model* for the risk of D.

Like RR, OR > 0, but unlike RR, OR *has no upper limit* whatever the baseline risk $P(D|\bar{E})$ is.

Thus, the OR can be effectively used as a scale for association even when $P(D|\bar{E})$ is large.

## Relative Effect

| Infant Mortality | Mother's Marital Status Unmarried | Married |
|---|---|---|
| Death | 16,712 | 18,784 |
| Live at 1 year | 1,197,142 | 2,878,421 |
| Total | 1,213,854 | 2,897,205 |

The OR for infant mortality associated with an unmarried mother is:

$$OR = \left[\frac{16712}{1213854} : \frac{1197142}{1213854}\right] : \left[\frac{18784}{2897205} : \frac{2878421}{2897205}\right] = 2.14$$

Associated with low birthweight, the OR is: $OR = \left[\frac{21054}{292323} : \frac{271269}{292323}\right] : \left[\frac{14442}{3818736} : \frac{3804294}{3818736}\right] = 20.4$

| Infant Mortality | Birthweight Low Birthweight | Normal Birthweight | Total |
|---|---|---|---|
| Death | 21,054 | 14,442 | 35,496 |
| Live at 1 year | 271,269 | 3,804,294 | 4,075,563 |
| Total | 292,323 | 3,818,736 | 4,111,059 |

# The odds ratio as an approximation to the relative risk

If the risk of disease is **low** - that is, the disease is **rare** - in both exposed and unexposed, $P(\bar{D}|E)$ and $P(\bar{D}|\bar{E})$ are both close to 1 and the OR and the RR are *approximately* equal:

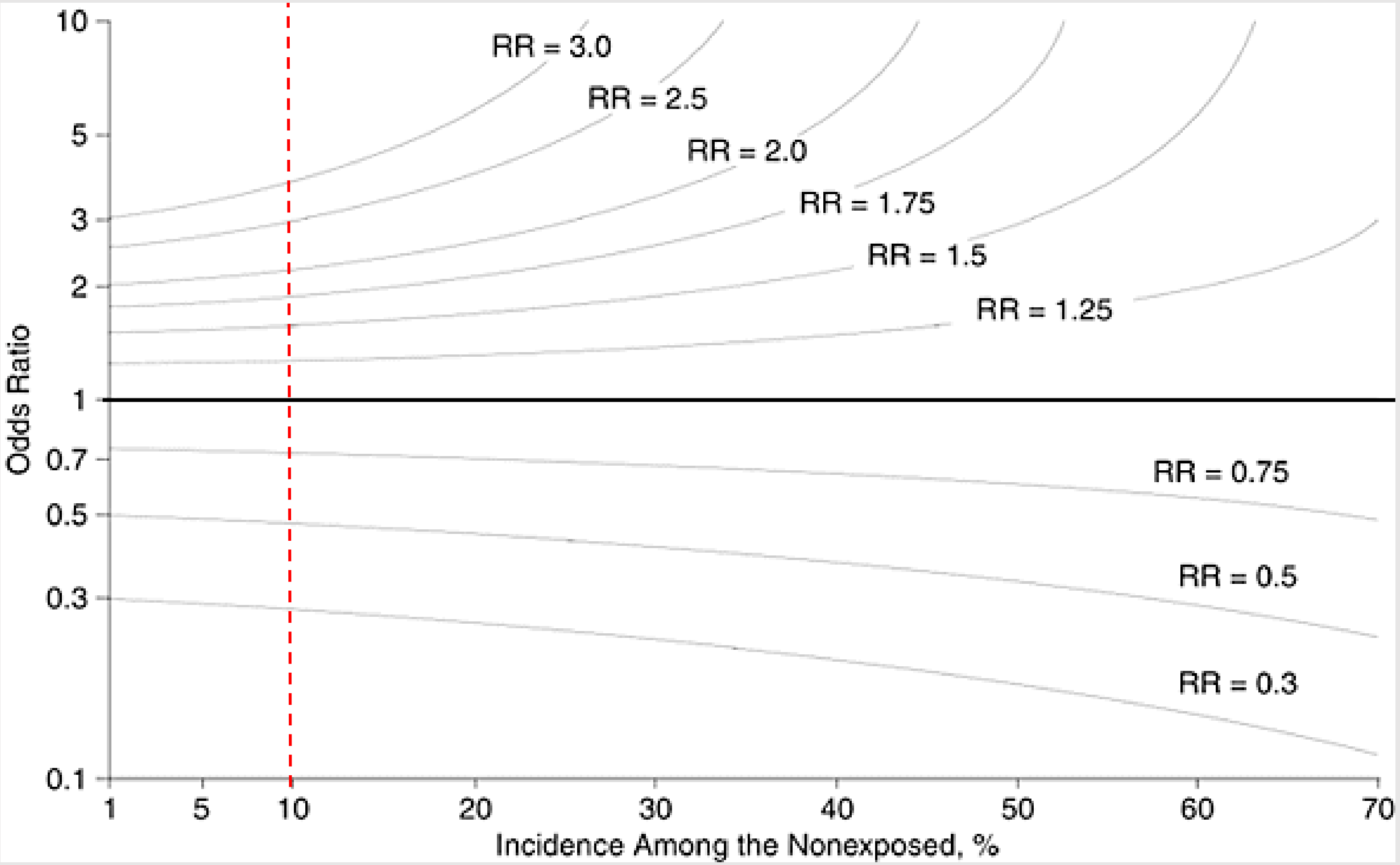
$$P(\bar{D}|E) \approx P(\bar{D}|\bar{E}) \approx 1 \qquad\qquad OR \approx \frac{P(D|E)}{P(D|\bar{E})} = RR$$

Generally, OR is similar to the RR when the *sum of the risks* - in the exposed and unexposed - is < 0.1

https://jamanetwork.com/journals/jama/fullarticle/188182



The relationship between relative risk (RR) and odds ratio (OR) by incidence of the outcome:




When the ***incidence*** of an outcome is low (<10%), the odds ratio is *close* to the relative risk.

The more frequent the outcome becomes, the more the odds ratio will **overestimate** the relative risk when it is more than 1 or **underestimate** the relative risk when it is less than 1.

https://www.youtube.com/watch?v=76W4Wymv2Ec

UNITÀ DI BIOSTATISTICA
Dipartimento Universitario Clinico di Scienze Mediche Chirurgiche e della Salute

# Symmetry of roles of disease and exposure in the odds ratio

The Odds Ratio is notoriously confusing when first encountered, particularly in contrast to the simplicity of the interpretation for the Relative Risk. Why is the Odds Ratio then used so often*? A fundamental reason is that the Odds Ratio is ***symmetric*** in the roles of D and E.

**Reversing** the roles of D and E makes **no difference** in Odds Ratio : this is the **key** to estimating association between an exposure and disease in **case-control studies** [**block 2**].

$$\boxed{OR = \frac{P(D|E)}{P(\bar{D}|E)} : \frac{P(D|\bar{E})}{P(\bar{D}|\bar{E})}} = \frac{P\,(D\&E)/P\,(E)}{P\,(\bar{D}\&E)/P\,(E)} : \frac{P\,(D\&\bar{E})/P\,(\bar{E})}{P\,(\bar{D}\&\bar{E})/P\,(\bar{E})}$$

$$= \frac{P(D\&E)}{P(\bar{D}\&E)} : \frac{P(D\&\bar{E})}{P(\bar{D}\&\bar{E})} = \frac{P(D\&E)}{P(D\&\bar{E})} : \frac{P(\bar{D}\&E)}{P(\bar{D}\&\bar{E})}$$

$$= \frac{P\,(D\&E)/P\,(D)}{P\,(D\&\bar{E})/P\,(D)} : \frac{P\,(\bar{D}\&E)/P\,(\bar{D})}{P\,(\bar{D}\&\bar{E})/P\,(\bar{D})}$$

$$\boxed{= \frac{P(E|D)}{P(\bar{E}|D)} : \frac{P(E|\bar{D})}{P(\bar{E}|\bar{D})}}$$

*Also for the popularity of the logistic regression model [**block 3**]…

UNITÀ DI BIOSTATISTICA
Dipartimento Universitario Clinico di Scienze Mediche Chirurgiche e della Salute

# Excess risk



To convey an **absolute** measure of the impact of exposure on risk, the Excess Risk, denoted by ER, could be estimated:

$$ER = P(D|E) - P(D|\bar{E})$$

The Excess Risk uses the same basic components as the Relative Risk (and the Odds Ratio), but looks at the **absolute**, rather than relative, difference in risk levels.

The Relative Risk for lung cancer associated with cigarette smoking is about **5** times as great as the Relative Risk for CHD due to smoking.

On the other hand, the Excess Risk for CHD **is larger** since it is the **most common** disease.

Therefore, from a health policy or public health point of view, cigarette intervention programs may be more important in terms of their **impact** on CHD.

## Absolute effect

$$ER \begin{cases} <0 & \text{a greater risk for D when unexposed than when exposed} \\ =0 & \text{independence of D and E} \\ >0 & \text{a greater risk for D when exposed than when unexposed} \end{cases}$$

$$-1 \leq \mathrm{ER} \leq 1$$

Excess Risk is the basis of an ***additive*** *model* for risk:

$$Risk_{Exposed} = Risk_{unexposed} + ER$$

Interpretation of the Excess Risk : difference in the number of cases in populations where either ***everyone*** is exposed or unexposed

Block 1.5

$P(D) = P(D|\bar{E})$ All **not exposed**: number of cases -> $\#cases = N * P(D|\bar{E})$

$P(D) = P(D|E)$ All **exposed**: number of cases -> $\#cases = N * P(D|E)$

Excess Risk : the “excess” number of cases when population members are *all exposed* as compared to them *all being unexposed*.

**Example:** study on the association between appendectomy and infections

Cumulative Incidence (CI) with appendectomy = 5.3% = 53/1000

Cumulative Incidence (CI) without appendectomy = 1.3% = 13/1000

Risk Difference (ER) = 40/1000= 4/100

**Interpretation**: Subjects who had an incidental appendectomy had 4 ***additional cases*** of wound infection per 100 people compared to subjects who did not have an incidental appendectomy. There were 4 *excess wound infections* per 100 subjects in the group that had incidental appendectomies, compared to the group without incidental appendectomy.

## Absolute effect

Excess Risk for infant mortality in the U.S. in 1991 associated with the mother's marital status:

| | Mother's Marital Status | |
|---|---|---|
| Infant Mortality | Unmarried | Married |
| Death | 16,712 | 18,784 |
| Live at 1 year | 1,197,142 | 2,878,421 |
| Total | 1,213,854 | 2,897,205 |

$$ER = \frac{16712}{1213854} - \frac{18784}{2897205} = 0.0073$$

Excess Risk for infant mortality associated with low birthweight:

| | Birthweight | | |
|---|---|---|---|
| Infant Mortality | Low Birthweight | Normal Birthweight | Total |
| Death | 21,054 | 14,442 | 35,496 |
| Live at 1 year | 271,269 | 3,804,294 | 4,075,563 |
| Total | 292,323 | 3,818,736 | 4,111,059 |

$$ER = \frac{21054}{292323} - \frac{14442}{3818736} = 0.0682$$

Low birthweight is *more influential* than marital status on both the absolute and relative comparative scales.

We would expect the infant mortality to increase by 7% if all births exhibited low birthweight as compared to all those being of normal birthweight (w.r.t 0.7% in case of marital status).

# Relative or absolute risk measures ?

Relative measures, such as relative risk, lose information on risk levels, so you can find relative risks *relatively low* associated with *very high absolute* differences, and viceversa.

Relative and absolute risk measures between incidence rates (per 100.000 pyrs) of disease in smokers and non-smokers:

| | Smokers | Not Smokers | RR | ER |
|---|---|---|---|---|
| Lung cancer | 48.33 | 4.49 | **10.8** | 43.84 |
| Cardiovascular disease | 294.67 | 169.54 | 1.7 | **125.13** |

For this reason, it is important to estimate in public health studies also the absolute differences between risks / rates (this is possible *in some types of studies* but not in others, block 2).

UNITÀ DI BIOSTATISTICA
Dipartimento Universitario Clinico di Scienze Mediche Chirurgiche e della Salute

For **uncommon** events such as clinically problematic rare adverse events, relative measures will tend to *exaggerate* differences. For **common** events such as therapeutic response, relative measures may *minimize* differences.

Knowledge of the **baseline rates** of the outcome of interest can help understand situations when the absolute difference is very small but the relative effect is very large.

Another possibility is to compute the so-called ***attributable risk*** measures that combine some of the advantages of both absolute and relative measures.

# Attributable risk

An individual may become diseased without being exposed to the risk factor of interest, that is $P(D|\bar{E}) \geq 0$.

Since in that scenario not **all** disease can be due to exposure, it is appealing to ask **how much** of the disease *D* in the population can be explained by the presence of the risk factor *E*.

The **Attributable Risk** is a measure of association designed to provide an answer to this question and is defined as **the fraction of all cases of D in the population (size N) that can be attributed to E.**

$$AR = \frac{N * P(D) - N * P(D|\bar{E})}{N * P(D)}$$

$$AR = \frac{P(D) - P(D|\bar{E})}{P(D)}$$

# Attributable risk

It could be demonstrated that: $$AR = \frac{P(E)[RR-1]}{1+P(E)[RR-1]}$$

Attributable Risk depends on the **strength** of the association between D and E (RR) and the **prevalence** of the risk factor E.

Therefore, it incorporates the advantages of both a relative and an absolute measure of association.

$$AR \begin{cases} < 0 & \text{exposure to } E \text{ is protective} \\ = 0 & \text{independence of D and E} \\ > 0 & \text{exposure to } E \text{ raises the risk of D} \end{cases}$$

$$-\infty < \mathrm{AR} \leq 1$$

AR can be an arbitrarily large negative number as the disease frequency becomes increasingly smaller and E is protective

The attractiveness of the AR is the insight it promises into the **potential impact** of an intervention program designed to **reduce exposure** to a risk factor E.

However, the assumption that the risk in the unexposed can be applied to individuals who are "changed" from E to not-E *through an intervention program* assumes essentially that the E–D relationship is **causal**.

An additional tacit assumption is that *modification* of an individual's E status does not alter **other risk factors**; in the extreme it is possible that reducing exposure to E may actually increase exposure to other risk factors and thereby make the disease burden greater.

For example, automobile drivers might respond to seat-belt laws by increasing their average speed, under a perception of increased safety, thereby offsetting mortality reductions introduced by higher seat-belt usage.

Block 1.5

Both of these concerns - ***causality*** and the ***effect*** of other factors - also apply to the RR and OR !!

[we will discuss the estimation of ***causal effects*** taking into account confounders either by design or using regression approaches, block 2/3]

| | Mother's Marital Status | |
|---|---|---|
| Infant Mortality | Unmarried | Married |
| Death | 16,712 | 18,784 |
| Live at 1 year | 1,197,142 | 2,878,421 |
| Total | 1,213,854 | 2,897,205 |

Attributable risk for marital status:

$$AR = \frac{0.0086 - 0.0065}{0.0086} = 0.25$$

| | Birthweight | | |
|---|---|---|---|
| Infant Mortality | Low Birthweight | Normal Birthweight | Total |
| Death | 21,054 | 14,442 | 35,496 |
| Live at 1 year | 271,269 | 3,804,294 | 4,075,563 |
| Total | 292,323 | 3,818,736 | 4,111,059 |

Attributable risk for low birthweight:

$$AR = \frac{0.0086 - 0.0038}{0.0086} = 0.56$$

UNITÀ DI BIOSTATISTICA
Dipartimento Universitario Clinico di Scienze Mediche Chirurgiche e della Salute

*Naive* interpretation : infant mortality could be reduced by 25% if all mothers were married, or by 56% if we could eliminate low birthweight infants.

While it is plausible that a substantial fraction of infant mortality could be prevented by intervention programs designed to eliminate the risk of a low birthweight child, it is not believable that 25% of infant deaths could be eradicated through a program to have single pregnant women marry before they give birth...

This suggests that marital status does not, in fact, **cause** infant mortality; the ***apparent*** association, as captured by either the Relative Risk, Odds Ratio, or Attributable Risk, is likely due to the effect of **other factors** that are related to both marital status and infant mortality.

One drawback in interpreting the AR is that it does not behave as a conventional fraction when more than one risk factor is examined.

That is, the AR for **two distinct** exposures **cannot be added** to give the AR for both factors considered simultaneously, even when the exposures are independently distributed.

| **Infant Mortality** | $E\ \&\ F$ | $E\ \&\ \bar{F}$ | $\bar{E}\ \&\ F$ | $\bar{E}\ \&\ \bar{F}$ | **Tot** |
|---|---|---|---|---|---|
| Death | 25497 | 5561 | 4084 | 354 | 35496 |
| Live at 1 yr | 1,002,268 | 1,022,204 | 1,023,681 | 1,027,410 | 4,075,563 |
| Tot | 1,027,765 | 1,027,765 | 1,027,765 | 1,027,764 | 4,111,059 |

Hypothetical data on two binary exposures, E and F, that might have generated the infant mortality data (the data have been set up so that *E* and *F* are independent)

**Block 1.5**

$P(D|E)$=(25497+5561)/(1027765+1027765)= 0.0151

$P(D|\bar{E})$=(4084+354)/(1027765+1027764)= 0.0022

$RR_E$=0.0151/0.0022=7 $AR_E$=0.75

$P(D|F)$=(4084+25497)/(1027765+1027765)=0.0144

$P(D|\bar{F})$=(5561+354)/(1027765+1027764)= 0.0029

$RR_F$=0.0144/0.0029=5 $AR_F$=0.67

$P(E) = P(F) = 0.5$

it appears as if infant death is 75% due to E; the other 67% is due to F ….

These two factors are independent and certainly the AR for both combined **cannot be the sum** of the individual ARs since this would greatly exceed 1 …

From another point of view, establishing the AR associated with E to be 0.75 cannot be interpreted as claiming that only 25% of infant mortality remains to be explained in the sense that AR for ***other factors*** will be 0.25 or smaller*.

Di Maso et al., Attributable fraction for multiple risk factors: Methods, interpretations, and examples, Stat Methods Med Res. 2019
https://pubmed.ncbi.nlm.nih.gov/31074326/

# Attributable risk in the exposed

The fraction attributable **in the exposed** is the proportion of cases attributable to exposure in the exposed population (i.e. when considering the only population on which exposure *can act*).

$$AR_{Exposed} = \frac{P(D|E) - P(D|\bar{E})}{P(D|E)}$$

It is the risk fraction of those exposed that is attributable to exposure.

$$AR_{Exposed} = \frac{RR - 1}{RR}$$

Note that we lose here the weight given by **prevalence** of the exposure in the population

This excess fraction represents the proportion of cases among the exposed that can be attributed to the exposure (*assuming causality*). In other words, it represents the proportion of cases among the exposed that ***could have been prevented*** if they had never been exposed.

Block 1.5

*Example 5.14. A total of 34 439 British male doctors were followed up for 40 years and their mortality in relation to smoking habits was assessed (Doll et al., 1994a). Mortality from certain diseases is shown in Table 5.3.*

| Underlying cause of death | Never smoked regularly Rate[b] (1) | Current cigarette smoker Rate[b] (2) | Rate ratio (2)/(1) | Rate difference[b] (2)–(1) | Excess fraction (%) $\frac{(2)-(1)}{(2)} \times 100$ |
|---|---|---|---|---|---|
| Cancer | | | | | |
| All sites | 305 | 656 | 2.2 | 351 | 54 |
| Lung | 14 | 209 | 14.9 | $\boxed{195}$ | 93 |
| Oesophagus | 4 | 30 | 7.5 | 26 | 87 |
| Bladder | 13 | 30 | 2.3 | 17 | 57 |
| Respiratory diseases (except cancer) | 107 | 313 | 2.9 | 206 | 66 |
| Vascular diseases | 1037 | 1643 | 1.6 | $\boxed{606}$ | 37 |
| All causes | 1706 | 3038 | 1.8 | 1332 | 44 |

[a] Data from Doll *et al.*, 1994a.
[b] Age-adjusted rates per 100 000 pyrs.

**44%** of deaths among male British doctors who smoked could be attributed to smoking (*assuming causality*).

The % of deaths that could be attributed to smoking varied by disease.

This % >> for lung cancer (93%) and << for vascular diseases (37%).

However, ***if smokers had never smoked***, the total # of deaths ***prevented*** >> for vascular diseases (606 per 100.000 pyrs) than for lung cancer (195 per 100.000 pyrs)

Therefore [again] here we have a difference between *AR in the exposed* and the absolute measures

UNITÀ DI BIOSTATISTICA
Dipartimento Universitario Clinico di Scienze Mediche Chirurgiche e della Salute

**Block 1.5**

Similar measures can be calculated when those exposed have **a lower risk** of developing the disease than those unexposed.

In these circumstances, we would have:

**Risk reduction:** $P(D|\bar{E}) - P(D|E)$

**Prevented fraction:** $\frac{P(D|\bar{E}) - P(D|E)}{P(D|\bar{E})}$

***Example 5.15.*** *Suppose that a group of oral contraceptive users and a group of never users were followed up in time and their ovarian cancer incidence was measured and compared. The results from this hypothetical study are shown in Table 5.4.*

| | Oral contraceptive use | |
|---|---|---|
| | Ever | Never |
| Ovarian cancer cases | 29 | 45 |
| Person-years at risk | 345 000 | 321 429 |
| Rate per 100 000 pyrs | 8.4 | 14.0 |

Rate ratio = 8.4 per 100 000 pyrs/14.0 per 100 000 pyrs = 0.60
Risk reduction = 14.0 per 100 000 pyrs – 8.4 per 100 000 pyrs = 5.6 per 100 000 pyrs.
Prevented fraction (%) = 100 × (5.6 per 100 000 pyrs / 14.0 per 100 000 pyrs) = 40%.

40% of ovarian cancer cases ***could have been prevented*** among never-users if they had used oral contraceptives